DOCUMENT RESUME

ED 373 091 TM 021 980

AUTHOR Wang, Chen-Shih; Ackerman, Terry
TITLE An Examination of Response Dependency When There Is More than One Correct Answer.
PUB DATE Apr 94
NOTE 18p.; Paper presented at the Annual Meeting of the American Educational Research Association (New Orleans, LA, April 4-8, 1994).
PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Analysis of Covariance; Elementary Education; Elementary School Students; Grade 3; Grade 6; Grade 8; Inferences; *Multiple Choice Tests; *Reading Tests; *Responses; Test Format; *Test Items; Thinking Skills
IDENTIFIERS *Dependence (Test Items); Expository Text; Illinois; *Illinois Goal Assessment Program; Local Independence (Tests); Narrative Text; Testlets

ABSTRACT

Passages used in the Illinois Goal Assessment Program (IGAP) reading test are intact pieces of literature, stories, and essays that match classroom reading assignments and typical student reading experiences. There are 15 testlets, each containing 5 items, associated with each passage. Each testlet requires students to demonstrate various levels of cognitive skills, from responding explicitly to drawing conclusions not directly stated, solving problems not discussed within the text, and using information derived from the passage. Because the texts often support more than one correct inference, the Illinois reading assessment uses a multiple-response (or multiple-correct) rather than multiple-choice format. This study investigates the magnitude of the dependence of items within and between testlets for IGAP narrative and expository subtests at grades 3, 6, and 8 (4,837, 4,840, and 5,011 examinees, respectively). Results suggest that the format has some effect on the issue of dependence, especially in the lower grades. Some degree of local dependence was found. It was larger within covariances than between covariances at all grades. One table and three figures present study findings. (Contains 5 references.) (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 373 091

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☒ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

TERRY ACKERMAN

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)"

An Examination of Response Dependency When There Is More Than One Correct Answer

Chen-Shih Wang
Terry Ackerman
University of Illinois at Urbana-Champaign

Running Head: Response Dependency

Paper presented at the 1994 Annual Meeting of the American Educational Research Association, New Orleans

TM021980

Abstract

In contrast to more traditional reading assessments that use isolated paragraphs and fragmented text, the passages used in the Illinois Goal Assessment Program (IGAP) are intact pieces of literature, stories, and essays that match classroom reading assignments and typical student reading experiences. There are 15 testlets, each containing five items, associated with each passage. Each testlet requires students to demonstrate various levels of cognitive skills, from explicit response to drawing conclusions that are not directly stated, solving problems not discussed within the text, and using information derived from the reading passage. Because texts often support more than one correct inference, the Illinois reading assessment uses a multiple response (or multiple correct) rather than a multiple choice format. The focus of this study is to investigate the dependency that results from such an assessment format. Specifically, this paper examines the magnitude of dependence of items within and between testlets for both IGAP subtests (narrative and expository) at grades three, six, and eight.

Introduction

In 1985 the Illinois State Legislature amended the school code to, among other things, identify and assess learning in six fundamental learning areas. Responding to the legislation, the Illinois State Board of Education directed committees comprised of educators throughout the state to establish broad goal statements in each of the fundamental areas. In 1989, the language arts subcommittee established the goal for reading which stated that schools need to prepare students to adequately read, comprehend, interpret, evaluate, and use written material.

For the purpose of assessment, reading was separated from the larger fundamental learning area of language arts. Beginning in the spring of academic year 1987-88, the Illinois Goal Assessment Program (IGAP) measured reading proficiency in grades three, six, and eight. Assessment of eleventh grade reading attainment was added in the spring of academic year 1989-90. The most recent assessment schedule began academic year 1992-93 where third, sixth, eighth, and tenth grade students were assessed in the spring of the year.

IGAP viewed reading as a dynamic process by which readers combine their background knowledge, their reading ability, their strategic awareness, and information supplied from a written text or passage to construct meaning. The format used to assess reading ability requires that the examinee read two genres of passages, one narrative (story

type) and one expository (informational type). Each passage is followed by 15 testlets about the passage. The test is administered in two 40-minute sessions with a minimum rest period of ten minutes between sessions.

Each IGAP reading testlets consists an item stem and five statements or inferences about the passage. The examinees are instructed that there may be one, two, or three correct inferences for each test item. In 1993 a new equating approach using item response theory (IRT) methodology was investigate. Because IRT assumes local independence, the amount of dependency created by the testlet format and multiple correct items was of major concern. Local independence, implies that when the abilities influencing test performance are held constant, examinees' responses to any pair of items be statistically independent. In other words, no relationship exists between examinees' responses to different items due to influence of the rule about number of correct inferences, or inference caused by being linked to the same stem and/or passage. Simply put, the response to each question is independent of or does not influence the response to any other item. (Hambleton, Swaminathan, and Rogers, 1991)

The purpose of this paper is to examine the issue of dependency between items. More specifically, the goal is to examine the dependency among the five items *within* each testlet as compared to items *across* testlets. Additionally, this paper investigates the effect of different item formats (one, two, or three correct inferences) on item

dependency? That is, are students more likely to follow the rule and mark the last two inferences as incorrect after they have marked the preceding three options as correct?

Method

Data sets of 4837, 4840, 5011 randomly selected examinees were obtained from grades three, six, and eight, respectively. The data set consisted of response patterns for five Yes-No items for each of the fifteen testlets in the narrative and the fifteen expository testlets. In sum, there was a total of 150 item responses per student. Each of the 150 responses was scored dichotomously (e.g., 1=right, 0=wrong) with the highest total subtest score 75 for both subtests.

Dependency was examined using two statistical methods. The first method involved comparing the average conditional covariances of the five items *within* a testlet to the average conditional covariances of items *between* testlets. The second method involved the construction of an item conditional empirical item characteristic curves (ICCs). Using this method the empirical ICC was determined for selected items conditioned on the total number correct score (with that item's score removed) and the number of YES answers to other items within the testlet.

FORTRAN programs were constructed to compute covariances and the ICCs for items within the narrative and expository subtests at each grade level.

Results

The analysis of the conditional covariance structure was computed to examine the degree of dependency *within* and *between* test items. For the simplicity of illustrating results the multiple correct answer format, mean *within* cluster covariance of 1-correct, 2-correct, 3-correct answers and a mean *between* cluster covariance conditioning on the total score excluding the testlet score (for *within* covariances) or testlet scores (for *between* comparisons). The mean covariances are calculated for both subtests (narrative and expository) across the three grade levels. The results are displayed in Table 1.

To provide a basis of comparison independent response data was generated using a 2PL IRT model. The mean covariance for this independent data resulted in a value of .0015.

Insert Table 1 about here

For all grade levels, the average *within* covariances (Table 1) are greater than the average *between* covariances for both genres. The average *between* covariance was about twice as large for the expository item pairs than the narrative passage item pair. In all cases the average

between covariances exceeded the value for the generated independent data.

This implies that there is more dependence within an item than between items. However, the magnitude of this dependency and whether IRT modeling is robust enough to this violation of local independence need to be studied further.

The mean *within* covariances of one-YES and three-YES clusters are generally greater than that of two-YES cluster. This is observed in the narrative subtests at all three grade levels. This suggests that the raw number correct score reflects not only the student's reading ability but also the ability of taking advantage of the rule (i.e. the multiple correct answers format). In the case of the three-YES cluster, students have a greater tendency to mark two NO responses if they have already marked 3 YES responses. Conversely, students tend to respond YES if they have responded four NO to the previous four choices. One may presume the students know there will be at least one YES response within the item. It seems that more dependence exists in the one-YES and three-YES clusters than in two-YES cluster.

Figures 1-3 below contain graphs of the conditional ICCs corresponding to the probabilities of examinees responding correctly to a selected item conditioned on the total test score (with the scores from the testlet containing the examined item excluded) **and** the number of YES responses that appear in the testlet held constant. This probability can be formulated as:

$$P(X=1| T, N_y)$$

where X represents correct response to one of the five selected parts, T is the total test (observed) score excluding the scores of the given item, and N_y is the number of YES responses the examinee has made within the examined item.

Insert Figure 1 about here

The ICCs in Figure 1 represents the probability of responding correctly (i.e., a NO response) to the fifth item of Testlet 14 in grade three narrative subtest. The correct pattern for this item is (Yes Yes Yes No No). In Plots 1-4, the dash-line ICCs represent the probabilities conditioning on the total scores **only** , the solid-lined ICCs represent conditioning on the total scores **and** number of YES responses: zero, one, two, and three. It can be seen in Plot one, that for those students who have no previous YES responses, the higher the total scores, the lower the probability of correctly responding to the fifth choice. That high-ability students tend to mark the fifth choice as "YES", suggests that high-ability students are more likely to take advantage of the rule than low-ability students. In contrast to Plot 1, the solid-lined ICCs of Plots 2-4 are less fluctuating. As the total score increases, the probabilities of marking the fifth choice as "NO" increase in the one-, two-, and three-YES cases. A somewhat dramatic situation occurs in three-YES case: the solid ICC is

very flat and straight indicating high probability of selecting NO for the fifth choice given the student has selected three previous YES responses.

In Figure 2, the ICCs of Item 1 in the grade six narrative subtest illustrate the probabilities of getting the second item correct when conditioning on the total scores and the number of YES responses.

Insert Figure 2 about here

The correct response pattern for this is (No Yes No Yes No). Notice that this two-YES test, the solid ICCs of Plot 1 fluctuate between probability range of 0.75 to 1.00 and ICCs for Plot 2 move up as the total scores increase. In contrast, the solid ICC of Plot 3 fluctuates mostly between 0.25 and 0.60 and the solid ICC of Plot 4 fluctuates between 0.0 and 0.25 with the tendency of decrease as ability or total score increases. That is, the higher the total scores(i.e., reading abilities), the less likely the student is to mark three YES's in this two-YES item.

Insert Figure 3 about here

The ICCs of item 11 in the grade eight Narrative subtest illustrated in Figure 3 are the probabilities of the examinees responding correctly to

the third item. The correct response pattern to this item is (No No Yes No No). For this one-YES correct answer case, most middle- to high-ability students have perfect chance to respond correctly on the third item if they make no YES response in the testlet. The solid ICCs of Plot 2 and Plot 3 fluctuate dramatically between probability range from .50 to 1.00. For Plot 4, While the solid ICC decreases at the grade six level as the total score increases (see Plot 4 of Figure 2), the solid ICC at the eighth grade level increases. In other words, even though the eighth grade students have marked YES to three other responses in the item, they still will answer with a fourth YES. This would imply that the eighth grade students, unlike third and sixth grade students, have a tendency to ignore the YES rule and concentrate on the item.

Discussion

Some degree of local dependence was found in the IGAP reading test data as suggested by the larger *within* covariances when compared to the *between* covariances at all grade levels for each genre.

Students' raw scores reflect not only their reading abilities but also their ability to take advantage of the rule. For third and sixth grade students, the greater the examinee's reading ability, the more likely the examinee is to take advantage of the rule. This did not appear to happen

as much for the eighth grade students. That is, students in the eighth grade appeared to be responding more to the question at hand, than to the rule. Thus, the unique multiple answer format seems to have some effect on the issue of dependence especially in the lower grades..

It remains to be determined as to the relationship between dependency and the robustness of IRT estimation procedure. Additionally, other factors, such as the stem effect, might confound with the effect of the multiple answer format, thereby influencing the dependence. For each item, the five statements or inferences are directly related to the question (the stem). When the examinees respond to each of the five choices, the response is affected by their reading ability, the rule, and the stem. More work needs to be done to determine how to separate and measure the influence of each of these components. Also, developing an appropriate algorithm and statistic for detecting the local dependence would be an interesting topic for future research concerning IGAP reading tests.

Reference

Hambleton, R. K., Swaminathan, H., and Rogers, H. J. (1991). MMSS Fundamentals of Item Response Theory. Newbury Park, CA: Sage.

Illinois State Board of Education (1989). The Illinois reading assessment: Classroom connections. Springfield, IL: Illinois State Board of Education.

Illinois State Board of Education (1992). The Illinois goal assessment program information bulletin. Springfield, IL: Illinois State Board of Education.

Illinois State Board of Education (1992). The Illinois goal assessment program: Reading. Springfield, IL: Illinois State Board of Education.

Lord, F. M.(1980). Applications of item response theory to practical testing programs. Hiils dale, NJ: Erlbaum.

Table 1 Within and between covariances by Narrative and Expository sub-tests for grades 3, 6, and 8

grade 3	Number of YES's	Narrative	Expository
mean within cluster covariance	1	0.0176	0.0149
	2	0.0135	0.0163
	3	0.0191	0.0243
mean between cluster covariance		0.0022	0.0022

grade 6	Number of YES's	Narrative	Expository
mean within cluster covariance	1	0.0113	0.0180
	2	0.0098	0.0156
	3	0.0105	0.0149
mean between cluster covariance		0.0019	0.0022

grade 8	Number of YES's	Narrative	Expository
mean within cluster covariance	1	0.0094	0.0109
	2	0.0068	0.0081
	3	0.0106	0.0132
mean between cluster covariance		0.0013	0.0020

Figure Captions

Figure 1: ICCs of Item 14 conditioning on the total scores and the number of YES's in grade 3 Narrative sub-test

Figure 2: ICCs of item 16 conditioning on the total scores and the number of YES's in grade 6 Narrative sub-test

Figure 3: ICCs of Item 11 conditioning on the total scores and the number of YES's in grade 8 Narrative sub-test

Figure 1

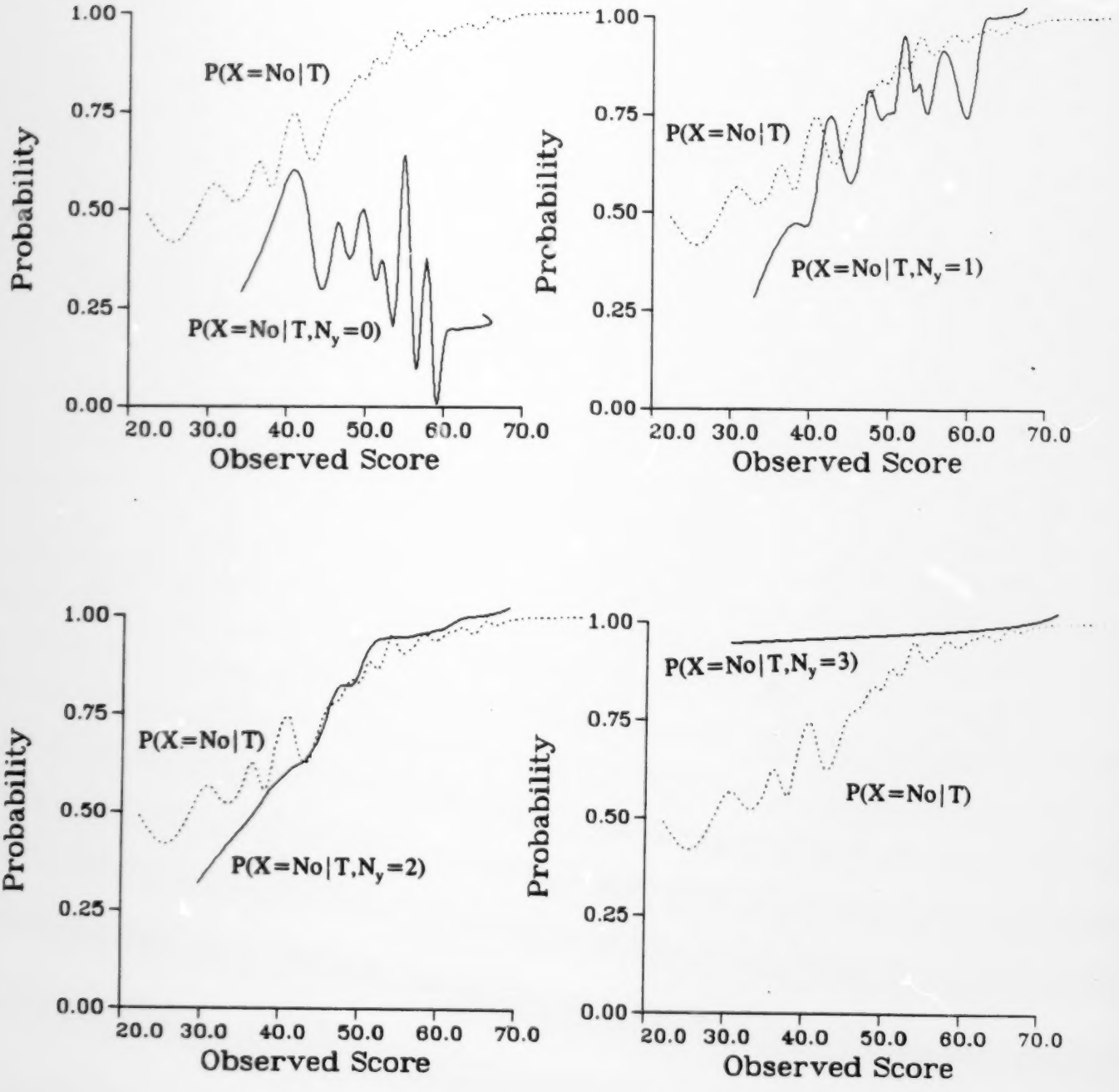

1.00
0.75
0.50
0.25
0.00
Probability
P(X=No|T)
P(X=No|T,N$_y$=0)
20.0 30.0 40.0 50.0 60.0 70.0
Observed Score
P(X=No|T)
P(X=No|T,N$_y$=1)
P(X=No|T,N$_y$=2)
P(X=No|T,N$_y$=3)

Figure 2

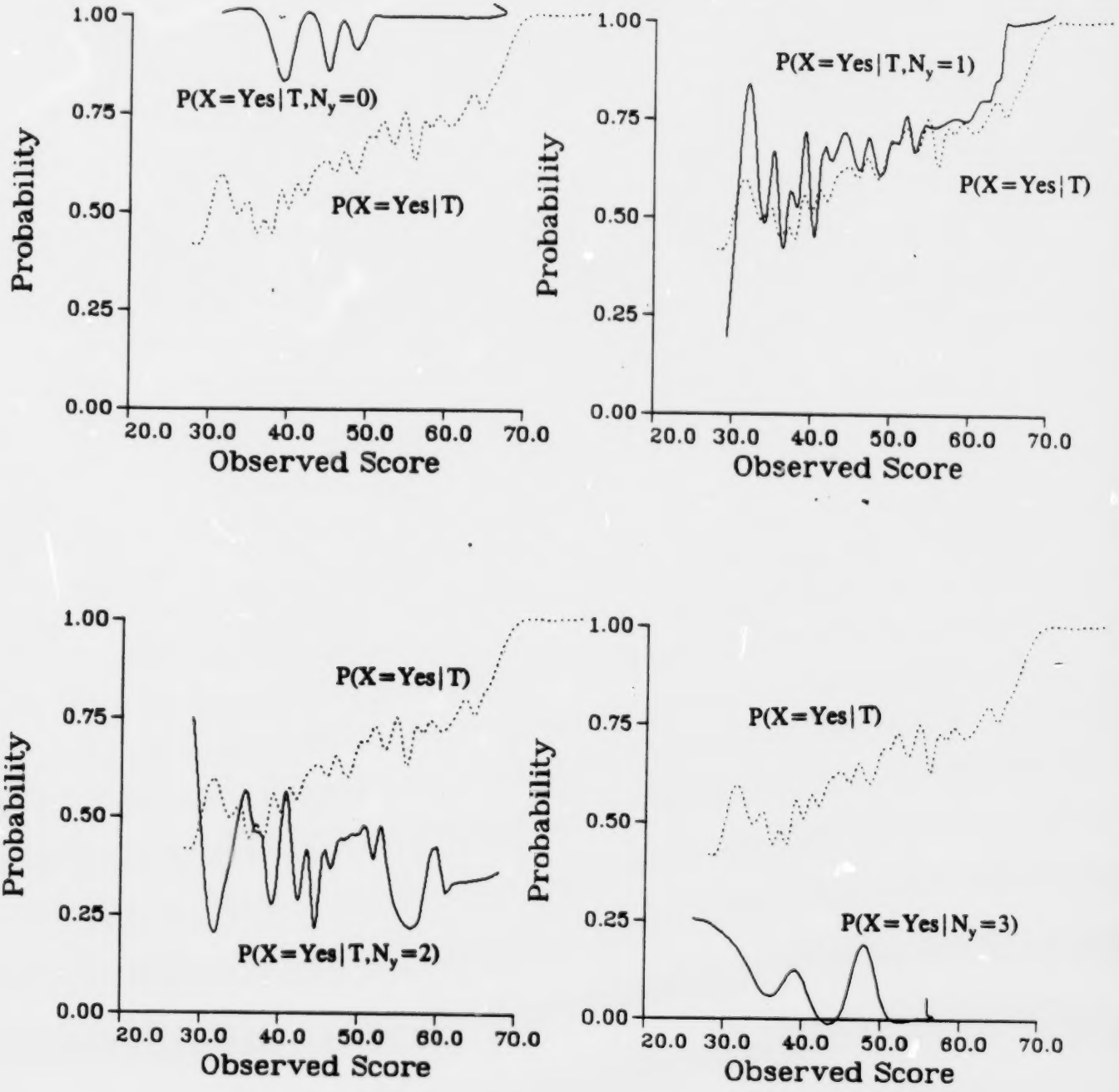
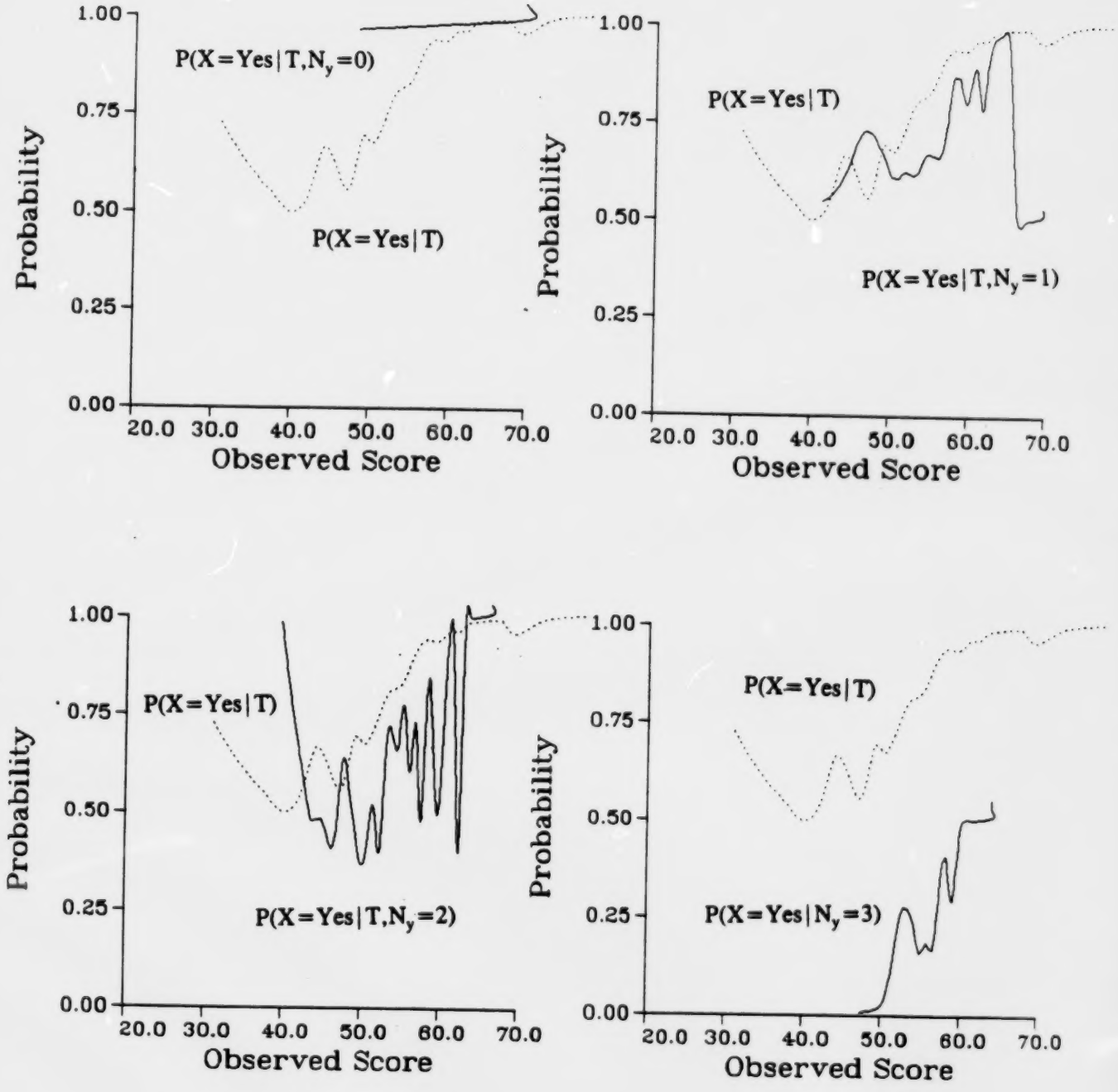

1.00
0.75
0.50
0.25
0.00
Probability
P(X=Yes|T,N$_y$=0)
P(X=Yes|T)
20.0 30.0 40.0 50.0 60.0 70.0
Observed Score
P(X=Yes|T,N$_y$=1)
P(X=Yes|T)
P(X=Yes|T)
P(X=Yes|T,N$_y$=2)
P(X=Yes|T)
P(X=Yes|N$_y$=3)

Figure 3


1.00
0.75
0.50
0.25
0.00
Probability
P(X=Yes|T,N_y=0)
P(X=Yes|T)
20.0 30.0 40.0 50.0 60.0 70.0
Observed Score
P(X=Yes|T)
P(X=Yes|T,N_y=1)
P(X=Yes|T)
P(X=Yes|T,N_y=2)
P(X=Yes|T)
P(X=Yes|N_y=3)